

This is a pre-copyedited, author-produced version of an article accepted for publication in Protein Engineering, Design and Selection, following peer review. The version of record J. Van Durme, G. De Baets, R. van der Kant, M. Ramakers, A. Ganesan, H. Wilkinson, R. Gallardo, F. Rousseau, and J. Schymkowitz, "Solubis: a webserver to reduce protein aggregation through mutation," Protein Eng Des Sel, p. gzw019, Jun. 2016. is available online at <https://academic.oup.com/peds/article/29/8/285/2223259?searchresult=1>.

Solubis: a webserver to reduce protein aggregation through mutation

Joost Van Durme^{1,2,*}, Greet De Baets^{1,2,*}, Rob Van der Kant^{1,2}, Meine Ramakers^{1,2}, Ashok Ganesan^{1,2}, Hannah Wilkinson^{1,2}, Rodrigo Gallardo^{1,2}, Frederic Rousseau^{1,2,*} and Joost Schymkowitz^{1,2,*}

¹ VIB Switch Laboratory, VIB, Leuven

² Switch Laboratory, Department of Cellular and Molecular Medicine, University of Leuven, Leuven, Belgium

⁺ Equally contributed to this study

* To whom correspondence should be addressed. Tel: +3216372570; Fax: +3216372751; Email: joost.schymkowitz@switch.vib-kuleuven.be. Correspondence may also be addressed to frederic.rousseau@switch.vib-kuleuven.be.

ABSTRACT

Protein aggregation is a major factor limiting the biotechnological and therapeutic application of many proteins, including enzymes and monoclonal antibodies. The molecular principles underlying aggregation are by now sufficiently understood to allow rational redesign of natural polypeptide sequences for decreased aggregation tendency, and hence potentially increased expression and solubility. Given that aggregation prone regions tend to contribute to the stability of the hydrophobic core or to functional sites of the protein, mutations in these regions have to be carefully selected in order not to disrupt protein structure or function. Therefore, we here provide access to an automated pipeline to identify mutations that reduce protein aggregation by reducing the intrinsic aggregation propensity of the sequence (using the TANGO algorithm), while taking care not to disrupt the thermodynamic stability of the native structure (using the empirical force field FoldX). Moreover, by providing a plot of the intrinsic aggregation propensity score of aggregation prone regions corrected by the local stability of that region in the folded structure, we allow users to prioritise those regions in the protein that are most in need of improvement through protein engineering. The method can be accessed at <http://solubis.switchlab.org/>.

KEYWORDS

Protein aggregation/Protein design/Structural bioinformatics

FUNDING

This work was supported by grants from the Flanders Institute for Biotechnology (VIB PRJ6], the University of Leuven [OT/12/092], the Funds for Scientific Research Flanders [G.0509.13], the Federal Office for Scientific Affairs of Belgium [IUAP P7/16], Boehringer Ingelheim Pharma GmbH & Co. KG to RVDK and by the European Research Council under the European Union's Horizon 2020 Framework Programme, ERC Grant agreement 647458 (MANGO) to JS.

INTRODUCTION

Protein aggregation is mediated by short aggregation prone regions (APRs), which assemble by intermolecular β -structured interactions that form the core of the aggregate. In native globular proteins these stretches are generally part of the hydrophobic core of the protein and hence protected from aggregation by the thermodynamic stability of the protein structure (Figure 1A). Besides structural stabilization a number of other mechanisms can also contribute to suppress aggregation (Balch *et al.*, 2008). One of them is the presence of aggregation gatekeeper residues, i.e. generally charged residues or prolines residues that slow down the aggregation reaction (Monsellier and Chiti, 2007, Otzen *et al.*, 2000, Richardson and Richardson, 2002, Rousseau *et al.*, 2006). In natural proteins sequences such gatekeeper residues are strongly enriched at the flanks of APRs. Moreover, molecular chaperones, such as Hsp70, bind to exposed APRs, preventing intermolecular assembly of APRs to nucleate aggregation (Van Durme *et al.*, 2009). Finally, protein turnover rates (De Baets *et al.*, 2011) and protein expression levels (Tartaglia *et al.*, 2009) are also tuned to minimize protein aggregation under physiological conditions. However, not all proteins are folded or adopt a globular conformation under physiological conditions to exert their function. These intrinsically disordered proteins (IDP) are characterized by a high structural flexibility but do perform key regulatory functions (i.e. signaling activities, DNA and RNA binding and cell cycle control). Therefore in IDPs the structural stabilization of APRs cannot significantly contribute to the reduction of protein aggregation. Therefore several specific sequence adaptations are present to maintain their solubility and prevent aggregation: IDPs have a high net charge, low hydrophobicity (Uversky, 2002), a lower number of APRs (Linding *et al.*, 2004), a higher proline content (Tomba, 2002) as well as the presence of so-called entropic bristle sequences (Santner *et al.*, 2012).

Aggregation of both globular proteins and IDPs have been associated with several pathologies, including neurodegenerative disorders such as Alzheimer's Disease (amyloid- β) and Parkinson's Disease (α -synuclein) (Ross and Poirier, 2004), as well as cancer (p53) (Xu *et al.*, 2011) and metabolic diseases (α -Galactosidase) (Siekierska *et al.*, 2012, Soong *et al.*, 2009). In the case of globular proteins, the aggregation problem is often exacerbated through mutations, which increase the solvent exposure of the aggregation prone regions by thermodynamically destabilizing the native structure (Dobson, 2004).

When globular proteins are employed for research, therapy or industrial applications, they need to withstand artificial conditions for which evolution has poorly equipped them. Given the ubiquitous nature of aggregation-prone sequences in the proteome, it is not surprising that protein aggregation is often observed when proteins are expressed far beyond their normal concentration in conditions with no or insufficient molecular chaperones. Moreover in technological applications, once purified, the proteins are expected to last far beyond their natural lifetime, allowing for critical nucleating events to start the protein aggregation reaction. Several methods have been developed to reduce protein aggregation of recombinant proteins, for example by using cell lines with increased chaperone content (Schlieker *et al.*, 2002), by generating fusion proteins with solubilizing tags (Park *et al.*, 2008,

Song *et al.*, 2011, Zhang *et al.*, 2004) or by careful formulation of buffers (Wang, 1999). Another approach would be to adapt the primary sequence to the new requirements through carefully selected mutations. We here present a webserver implementing a rational design strategy, called the Solubis (Ganesan *et al.*, 2016), which employs the FoldX (Schymkowitz *et al.*, 2005) and TANGO (Fernandez-Escamilla *et al.*, 2004) algorithms to identify selected mutations that render a protein less aggregation-prone, while maintaining or even improving its intrinsic stability and function (Ganesan, Siekierska, Beerten, Brams, Van Durme, De Baets, Van der Kant, Gallardo, Ramakers, Langenberg, Wilkinson, De Smet, Ulens, Rousseau and Schymkowitz, 2016).

RESULTS

The interplay between protein stability and intrinsic aggregation propensity

Over 80% of proteins possess at least one APR in their primary sequence (Rousseau *et al.*, 2006), but most often they form an integral part of the hydrophobic core of the protein (De Baets *et al.*, 2014) and are thus protected from higher order interactions. Notable exceptions are active site and protein-protein interaction sites, where functional constraints result in APRs in solvent exposed parts of the structure (Castillo and Ventura, 2009, Kumar *et al.*, 2011, Wang *et al.*, 2010). Therefore, when considering protein aggregation under native conditions, it is key to estimate the likelihood of each APR in the protein to become solvent exposed and thus available for aggregation with the identical sequence in other molecules. To achieve this, Trout *et al.* have proposed to correct the hydrophobicity of each amino acid by the fraction of surface that is solvent exposed (Voynov *et al.*, 2009). However, this method has two limitations: (i) it has long been known that beta-sheet propensity and local net charge are equally important determinants of local aggregation propensity (Chiti *et al.*, 2003) and (ii) the degree of solvent exposure in the crystal structure underestimates the exposure of APRs through local unfolding/dynamics (Chiti and Dobson, 2009). To address these issues, we use (1) the TANGO algorithm to identify genuine APRs and to score the intrinsic aggregation propensity of the unfolded protein sequence and (2) use the empirical all-atom force-field FoldX to estimate the contribution of each APR to the local stability of the protein (ΔG value) (Schymkowitz, Borg, Stricher, Nys, Rousseau and Serrano, 2005). We display the total aggregation propensity of a protein using so-called stretch plots, where each APR is represented as a single point in a two-dimensional space where the intrinsic aggregation propensity of the region is on the ordinate and its contribution to protein stability is on the abscissa (Figure 1B). In this plot, the APRs are more problematic as they move towards to upper right corner of the plot. In order to be able to rank the aggregation of protein variants we also developed the Solubis score which allows capturing the information of the stretch-plot by a single numeral. The Solubis score results from calculating the summed ΔG of the APR, which is cut off between -5 and 5 and rescaled to values between 0 and 1. For each residue, this value is multiplied by the TANGO score, resulting in the Solubis score. (Figure 1C). In this manner, the TANGO score of the most buried APRs is reduced, thereby eliminating these as determinants of aggregation nucleation under native conditions. As illustrated in the plots, the Protective Antigen protein of *Bacillus anthracis* has different APRs identified in its sequence (Figure 1A&B), but only one seems particularly important under native conditions according to Solubis (Figure 1C). The importance of this stretch has previously been demonstrated experimentally (Ganesan *et al.*, 2012). Moreover, it has also been experimentally demonstrated that mutations in this stretch lead to a reduction of the aggregation of Protective antigen can be identified using the Solubis approach by scanning the Mutational Aggregation and Stability Spectrum of this APR, Figure 1D, see below). Mutant PA containing such a mutation was demonstrated to display significantly decrease aggregation kinetics at mildly elevated temperatures (40 °C, Figure 1E) while retaining its native structure and function (Ganesan, Siekierska, Beerten, Brams, Van Durme, De Baets, Van der Kant, Gallardo, Ramakers, Langenberg, Wilkinson, De Smet, Ulens, Rousseau and Schymkowitz, 2016) .

The Solubis Pipeline

Protein aggregation nucleating regions can be identified using specialised software, which has been reviewed elsewhere (Belli *et al.*, 2011, De Baets *et al.*, 2014). We here employed the statistical thermodynamics algorithm TANGO (Fernandez-Escamilla, Rousseau, Schymkowitz and Serrano, 2004) to detect aggregation hotspots in the target sequence. This algorithm was demonstrated to have the highest specificity in a recent independent comparison of aggregation predictors, resulting in a low number of false positives (Tsolis *et al.*, 2013). For the purpose of guiding the design of mutations that will be experimentally verified, this is an attractive feature since it maximises the chance that mutations actually reduce the aggregation tendency. The downside of this choice is a relatively low sensitivity, compared to other predictors (e.g. Aggrescan (Conchillo-Sole *et al.*, 2007) or the metapredictor Amylpred2 (Tsolis, Papandreou, Ionomidou and Hamodrakas, 2013)), which can result in the failure to detect some APRs. Therefore, as protein solubility correlates with the number of APRs in a protein (Ganesan, Siekierska, Beerten, Brams, Van Durme, De Baets, Van der Kant, Gallardo, Ramakers, Langenberg, Wilkinson, De Smet, Ulens, Rousseau and Schymkowitz, 2016), reducing the aggregation propensity of APRs predicted with high specificity to contribute to aggregation under native conditions, is probably the most robust method to reduce aggregation.

As it is important to consider mutations that reduce intrinsic aggregation without thermodynamically destabilising the native structure, we select proteins for which high-resolution crystallographic structures are available. Moreover, this structural information also enables us to visualize the topological position of the aggregating regions using atomic structure viewers. Methods that have reasonable accuracy in predicting the mutational effects on protein stability have been reviewed elsewhere (Potapov *et al.*, 2009) and the results shown here were obtained exclusively with the FoldX forcefield (Schymkowitz, Borg, Stricher, Nys, Rousseau and Serrano, 2005), which shows excellent performance in protein redesign (Kiel and Serrano, 2014, McKeone *et al.*, 2014).

Two classes of mutations can be designed to reduce protein aggregation: (i) mutations that eliminate or strongly reduce the intrinsic aggregation propensity of the sequence, thereby slowing down the aggregation reaction and (ii) mutations that stabilise the interaction of the aggregating region with the rest of the structural domain in which it resides, thus providing additional protection from solvent exposure. In the ideal case, mutations can be identified that unify both goals, but often a combination of mutations is required to maximally suppress aggregation. Reduction of intrinsic aggregation is usually achieved by the introduction of aggregation breaking residues, called gatekeepers (Monsellier and Chiti, 2007, Rousseau, Serrano and Schymkowitz, 2006), in the aggregation nucleating sequences. Since the gatekeepers consist of the charged amino acids (Arg, Lys, Glu, Asp) and proline, most often they need to be placed in loop regions in order not to disturb the hydrophobic core of the protein. The Solubis method thus consists of systematically mutating the residues residing within an APR (or TANGO zone) to each of the gatekeeper residues and calculating the consequent change in TANGO score as well as the change in the thermodynamic stability of the protein using FoldX (this process will be called gatekeeper scan in what follows). The results of a computational gatekeeper scan of each APR can be displayed as a scatter plot (Figure 1D) of the change in

thermodynamic stability ($\Delta\Delta G$ values calculated by FoldX in kcal/mol) versus change in the intrinsic aggregation propensity. The latter values are calculated by TANGO and range between 0 and 100 per amino acid residue. In the plot the sum of TANGO scores of all residues in a given APR is given. These Mutant Aggregation & Stability Spectrum (MASS) plots (Siekierska *et al.*, 2012) allow easy identification of ideal mutations, which have large negative values on both axes. This approach has been experimentally validated by improving the solubility and abundance of both protective antigen and α -galactosidase (Ganesan, Siekierska, Beerten, Brams, Van Durme, De Baets, Van der Kant, Gallardo, Ramakers, Langenberg, Wilkinson, De Smet, Ulens, Rousseau and Schymkowitz, 2016). Moreover, a survey of published solubility-increasing variants shows that 75% of the variants lower the aggregation tendency, illustrating the relevance of this method (De Baets *et al.*, 2015).

Solubis database

Currently the database contains data on 74000 mutations analysed in 585 high-quality structures (sequence identity < 30%, R-factor < 0.19 and resolution < 1.5 Å) (Joosten *et al.*, 2011) on which the Solubis approach was run. The database interface allows users to search for mutations by filtering on UniProtID, PDB ID, mutated residue (P, R, K, D or E) and the phenotypic effects. These effects include changes in aggregation tendency ($\Delta TANGO$) and structural stability ($\Delta\Delta G$) upon mutation. Applying the filter settings results in a set of variants that fulfil the requirements.

Job Submission and results viewing

The Solubis website is free and open to all users and there is no login requirement. The execution time can range from a few minutes to about an hour. This depends on the number of APRs present in the protein, the type of selected gatekeepers and the size of the protein. In the following, a description of the required input and the output are given in more detail.

Input interface- In a first step, users need to define the structure they want to optimize either by providing a PDB ID or an uploaded PDB file. After selecting the chain they want to analyze; users can define the TANGO threshold i.e. the value above which a short stretch is treated as an aggregation prone region. All the defined APRs will be analyzed by the Solubis approach, meaning that the higher this threshold, only the stronger APRs will be used for a Solubis analysis. As a last step, users need to select the gatekeeper residues to mutate the TANGO zone residues to. You can choose to mutate to all gatekeeper residues (P, R, K, D and E) or select only specific ones.

Output- The Solubis output contains an overview of the identified APRs, which is represented by the above-described stretch plot (Figure 1B), where the intrinsic aggregation propensity of each APR is plotted against the contribution of this stretch to protein stability and a plot of the ΔG -rescaled TANGO aggregation propensity versus the primary sequence. Moreover, the aggregation-prone regions are also visualized in the original protein structure (Figure 1B). These plots illustrate in a clear way the location of aggregate prone regions in the protein and allow the user to identify the problematic APRs under native conditions.

For each of these APRs, a mutation scan to the selected gatekeepers is performed. The effect of these mutations on thermodynamic stability and intrinsic aggregation propensity is visualized in an above-described MASS-plot (Figure 1D) per APR, allowing the user to identify interesting mutations to modulate each relevant APR. We also offer a tab-delimited file containing an overview of all the mutations performed.

SUMMARY

Over 80% of the proteome has at least one aggregation prone region within its primary sequence. As a result, protein aggregation is often encountered when proteins are overexpressed or recombinantly produced. Moreover, aggregation represents a major liability with respect to the immunogenicity of biotherapeutics. However, redesigning globular proteins to eliminate aggregation is not a straightforward task, as most aggregation nucleating sequences are part of the hydrophobic core and therefore difficult to mutate without disrupting protein structure and function.

We developed a minimal redesign method, termed Solubis, to abrogate aggregation by silencing aggregation nucleating sequences with single point mutations, which are selected to maximally reduce the intrinsic aggregation propensity of the sequence whilst preserving thermodynamic stability of the functional protein. Our *in silico* method allows sifting hundreds to thousands of mutations, simultaneously evaluating protein aggregation and stability, typically detecting 1 to 5 appropriate aggregation-reducing mutations per target protein that are compatible with the native protein structure and stability.

FIGURE CAPTION

Figure 1: **Solubis analysis of the Protective Antigen protein from *Bacillus Anthracis*.** A) Visualization of the APRs identified by TANGO in the protein structure (PDBID: 1ACC). B) Stretch plot where each point represents the intrinsic aggregation propensity and contribution to protein stability of a specific APR. The upper right corner represent exposed APRs with a high aggregation tendency. C) Plot of the normal TANGO aggregation propensity (TANGO) and the ΔG -rescaled TANGO aggregation propensity (SolubiS) versus the primary sequence. D) Mutant Aggregation and Stability Spectrum (MASS) plot visualizing the effect of each mutation on the change in thermodynamic stability ($\Delta\Delta G$) versus the change in intrinsic aggregation propensity ($\Delta TANGO$). (E) Aggregation kinetics of protective antigen wild type (red) and solubis mutant T576E/S559L at 40°C measured by right angle light scattering at 475nm using the OPTIM1000 (Unchained labs) (Reproduced with permission from the authors from (Ganesan, Siekierska, Beerten, Brams, Van Durme, De Baets, Van der Kant, Gallardo, Ramakers, Langenberg, Wilkinson, De Smet, Ulens, Rousseau and Schymkowitz, 2016))

REFERENCES

- Balch W.E., Morimoto R.I., Dillin A. and Kelly J.W. (2008) Adapting proteostasis for disease intervention. *Science*, 319, 916-919.
- Belli M., Ramazzotti M. and Chiti F. (2011) Prediction of amyloid aggregation in vivo. *EMBO Rep*, 12, 657-663.
- Castillo V. and Ventura S. (2009) Amyloidogenic Regions and Interaction Surfaces Overlap in Globular Proteins Related to Conformational Diseases. *Plos Computational Biology*, 5.
- Chiti F. and Dobson C.M. (2009) Amyloid formation by globular proteins under native conditions. *Nat Chem Biol*, 5, 15-22.
- Chiti F., Stefani M., Taddei N., Ramponi G. and Dobson C.M. (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature*, 424, 805-808.
- Conchillo-Sole O., de Groot N.S., Aviles F.X., Vendrell J., Daura X. and Ventura S. (2007) AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics*, 8, 65.
- De Baets G., Reumers J., Delgado Blanco J., Dopazo J., Schymkowitz J. and Rousseau F. (2011) An evolutionary trade-off between protein turnover rate and protein aggregation favors a higher aggregation propensity in fast degrading proteins. *PLoS Comput Biol*, 7, e1002090.
- De Baets G., Schymkowitz J. and Rousseau F. (2014) Predicting aggregation-prone sequences in proteins. *Essays Biochem*, 56, 41-52.
- De Baets G., Van Durme J., Rousseau F. and Schymkowitz J. (2014) A genome-wide sequence-structure analysis suggests aggregation gatekeepers constitute an evolutionary constrained functional class. *J Mol Biol*.
- De Baets G., Van Durme J., van der Kant R., Schymkowitz J. and Rousseau F. (2015) Solubis: Optimize your protein. *Bioinformatics*, 31, 2580-2.
- Dobson C.M. (2004) Principles of protein folding, misfolding and aggregation. *Seminars in Cell & Developmental Biology*, 15, 3-16.
- Fernandez-Escamilla A.M., Rousseau F., Schymkowitz J. and Serrano L. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol*, 22, 1302-1306.
- Ganesan A., Siekierska A., Beerten J., Brams M., Van Durme J., De Baets G., Van der Kant R., Gallardo R., Ramakers M., Langenberg T. et al. (2016) Structural hot spots for the solubility of globular proteins. *Nature communications*, 7, 10816.
- Ganesan A., Watkinson A. and Moore B.D. (2012) Biophysical characterisation of thermal-induced precipitates of recombinant anthrax protective antigen: Evidence for kinetically trapped unfolding domains in solid-state. *European journal of pharmaceuticals and biopharmaceutics : official journal of Arbeitsgemeinschaft fur Pharmazeutische Verfahrenstechnik eV*.
- Joosten R.P., te Beek T.A., Krieger E., Hekkelman M.L., Hooft R.W., Schneider R., Sander C. and Vriend G. (2011) A series of PDB related databases for everyday needs. *Nucleic Acids Res*, 39, D411-419.
- Kiel C. and Serrano L. (2014) Structure-energy-based predictions and network modelling of RASopathy and cancer missense mutations. *Molecular systems biology*, 10, 727.
- Kumar S., Singh S.K., Wang X.L., Rup B. and Gill D. (2011) Coupling of Aggregation and Immunogenicity in Biotherapeutics: T- and B-Cell Immune Epitopes May Contain Aggregation-Prone Regions. *Pharmaceutical Research*, 28, 949-961.
- Linding R., Schymkowitz J., Rousseau F., Diella F. and Serrano L. (2004) A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *J Mol Biol*, 342, 345-353.
- McKeone R., Wikstrom M., Kiel C. and Rakoczy P.E. (2014) Assessing the correlation between mutant rhodopsin stability and the severity of retinitis pigmentosa. *Molecular vision*, 20, 183-199.
- Monsellier E. and Chiti F. (2007) Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO reports*, 8, 737-742.
- Otzen D.E., Kristensen O. and Oliveberg M. (2000) Designed protein tetramer zipped together with a hydrophobic Alzheimer homology: a structural clue to amyloid assembly. *Proc Natl Acad Sci U S A*, 97, 9907-9912.
- Park J.S., Han K.Y., Lee J.H., Song J.A., Ahn K.Y., Seo H.S., Sim S.J., Kim S.W. and Lee J. (2008) Solubility enhancement of aggregation-prone heterologous proteins by fusion expression using stress-responsive *Escherichia coli* protein, RpoS. *BMC biotechnology*, 8, 15.

Potapov V., Cohen M. and Schreiber G. (2009) Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Eng Des Sel*, 22, 553-560. First published on 2009/06/30, doi: 10.1093/protein/gzp030.

Richardson J.S. and Richardson D.C. (2002) Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc Natl Acad Sci U S A*, 99, 2754-2759.

Ross C.A. and Poirier M.A. (2004) Protein aggregation and neurodegenerative disease. *Nature medicine*, 10 Suppl, S10-17.

Rousseau F., Schymkowitz J. and Serrano L. (2006) Protein aggregation and amyloidosis: confusion of the kinds? *Curr Opin Struct Biol*, 16, 118-126.

Rousseau F., Serrano L. and Schymkowitz J.W. (2006) How evolutionary pressure against protein aggregation shaped chaperone specificity. *J Mol Biol*, 355, 1037-1047.

Santner A.A., Croy C.H., Vasanwala F.H., Uversky V.N., Van Y.Y. and Dunker A.K. (2012) Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression. *Biochemistry*, 51, 7250-7262.

Schlieker C., Bukau B. and Mogk A. (2002) Prevention and reversion of protein aggregation by molecular chaperones in the *E. coli* cytosol: implications for their applicability in biotechnology. *Journal of biotechnology*, 96, 13-21.

Schymkowitz J., Borg J., Stricher F., Nys R., Rousseau F. and Serrano L. (2005) The FoldX web server: an online force field. *Nucleic Acids Res*, 33, W382-388.

Siekierska A., De Baets G., Reumers J., Gallardo R., Rudyak S., Broersen K., Couceiro J., Van Durme J., Schymkowitz J. and Rousseau F. (2012) alpha-Galactosidase aggregation is a determinant of pharmacological chaperone efficacy on Fabry disease mutants. *J Biol Chem*.

Siekierska A., De Baets G., Reumers J., Gallardo R., Rudyak S., Broersen K., Couceiro J., Van Durme J., Schymkowitz J. and Rousseau F. (2012) alpha-Galactosidase aggregation is a determinant of pharmacological chaperone efficacy on Fabry disease mutants. *J Biol Chem*, 287, 28386-28397.

Song J.A., Lee D.S., Park J.S., Han K.Y. and Lee J. (2011) A novel *Escherichia coli* solubility enhancer protein for fusion expression of aggregation-prone heterologous proteins. *Enzyme and microbial technology*, 49, 124-130.

Soong R., Brender J.R., Macdonald P.M. and Ramamoorthy A. (2009) Association of highly compact type II diabetes related islet amyloid polypeptide intermediate species at physiological temperature revealed by diffusion NMR spectroscopy. *J Am Chem Soc*, 131, 7079-7085.

Tartaglia G.G., Pechmann S., Dobson C.M. and Vendruscolo M. (2009) A relationship between mRNA expression levels and protein solubility in *E. coli*. *J Mol Biol*, 388, 381-389.

Tompa P. (2002) Intrinsically unstructured proteins. *Trends Biochem Sci*, 27, 527-533.

Tsolis A.C., Papandreou N.C., Iconomidou V.A. and Hamodrakas S.J. (2013) A consensus method for the prediction of 'aggregation-prone' peptides in globular proteins. *PLoS One*, 8, e54175.

Uversky V.N. (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*, 11, 739-756. First published on.

Van Durme J., Maurer-Stroh S., Gallardo R., Wilkinson H., Rousseau F. and Schymkowitz J. (2009) Accurate prediction of DnaK-peptide binding via homology modelling and experimental data. *PLoS Comput Biol*, 5, e1000475.

Voynov V., Chennamsetty N., Kayser V., Helk B. and Trout B.L. (2009) Predictive tools for stabilization of therapeutic proteins. *Mabs*, 1, 580-582.

Wang W. (1999) Instability, stabilization, and formulation of liquid protein pharmaceuticals. *International journal of pharmaceuticals*, 185, 129-188. .

Wang X.L., Singh S.K. and Kumar S. (2010) Potential Aggregation-Prone Regions in Complementarity-Determining Regions of Antibodies and Their Contribution Towards Antigen Recognition: A Computational Analysis. *Pharmaceutical Research*, 27, 1512-1529.

Xu J., Reumers J., Couceiro J.R., De Smet F., Gallardo R., Rudyak S., Cornelis A., Rozenski J., Zwolinska A., Marine J.-C. et al. (2011) Gain of function of mutant p53 by coaggregation with multiple tumor suppressors. *Nature chemical biology*, 7, 285-295.

Zhang Y.B., Howitt J., McCorkle S., Lawrence P., Springer K. and Freimuth P. (2004) Protein aggregation during overexpression limited by peptide extensions with large net negative charge. *Protein expression and purification*, 36, 207-216.

Van Durme et al, Figure 1

